# The MoodStripe - An evaluation of a novel visual interface as an alternative for online response gathering

**Matevž Pesek,**
**Primož Godec,**
**Mojca Poredoš**
University of Ljubljana, Faculty of computer and information science
matevz.pesek@fri.uni-lj.si
{primoz.godec, mojca.poredos}
@lgm.fri.uni-lj.si

**Gregor Strle**
Scientific Research Centre of the Slovenian Academy of Sciences and Arts
gregor.strle@zrc-sazu.si

**Jože Guna**
**Emilija Stojmenova**
**Matevž Pogačnik**
University of Ljubljana, Faculty of electrotechnics
{joze.guna, emilija.stojmenova,
matevz.pogacnik}@fe.uni-lj.si

**Matija Marolt**
University of Ljubljana, Faculty of computer and information science
matija.marolt@fri.uni-lj.si

**ABSTRACT**
We present an innovative dynamic visual interface, the MoodStripe, which provides a continuous-scale, multi-parameter drag-and-drop alternative to the standard *n*-degree (Likert) scale widgets, commonly used in online evaluation processes. We elaborate on the motivation for the development of the new user input interfaces, and present the results of cross evaluation of the GMail product by using the SUS questionnaire with the standard and the proposed MoodStripe interfaces. The overall goal is to design a more intuitive interface, by reducing the noise and task load inherent in traditional interfaces for standardized user-feedback gathering tests. The results show the MoodStripe interface outperforms the standard scale approach both in terms of intuitiveness and functionality. Additionally, the cross-evaluation of the both approaches shows comparable SUS scores.

**Author Keywords**
user interface; gathering feedback efficiency; human computer interaction; system usability score; user evaluation study

**ACM Classification Keywords**
H.5.2. Information Interfaces and Presentation: User Interfaces

**INTRODUCTION**
The concept of user experience (UX) [1] has become one of the driving factors of interface design. There are several domains embracing the motivation for such research, including human-computer interaction (HCI), gaming, business and social applications, marketing, and others. Common to all is the iterative process of user centered design, of building up user experience by evaluating the product, typically by independent users or testers, modifying the initial prototype and incorporating the results into the next version of the product. There are several ways of gathering user feedback, including interviews, observations, printed handouts and on-line surveys, among others. We believe that the type of user interface selected can affect the feedback gathering procedure and in some cases influence (or bias) the results. In contrast to the evaluated interface, we focus our study on the UX during the evaluation procedure. Moreover, we also devote our attention to the quality of gathered responses.

In this paper we focus on on-line survey evaluation forms, one of the most commonly used media for user-feedback gathering. There are several metrics available for such evaluation, among most popular being NASA TLX [4], SUS [2], UMUX [6] and AttrakDiff [5]. The questionnaires are commonly implemented as on-line web pages, using standard HTML widgets, e.g. radio buttons, check boxes and text fields. In the past, such implementation was necessary to avoid the compatibility issues between different platforms and web browsers. However, recent development has brought towards unification of supported technologies, with JavaScript language becoming the standard for constructing dynamic interfaces.

In what follows, we first discuss the advantages offered by the proposed dynamic interface MoodStripe in comparison to a widely accepted standard *n*-degree (Likert) scale. Both approaches were subjected to the system usability score (SUS) questionnaire and evaluated in terms of user-friendliness, intuitiveness and functionality. The efficiency of the MoodStripe was further evaluated in terms of consistency by measuring the noise and comparing it to the *n*-degree scale as the golden standard. The interface evaluation procedure was

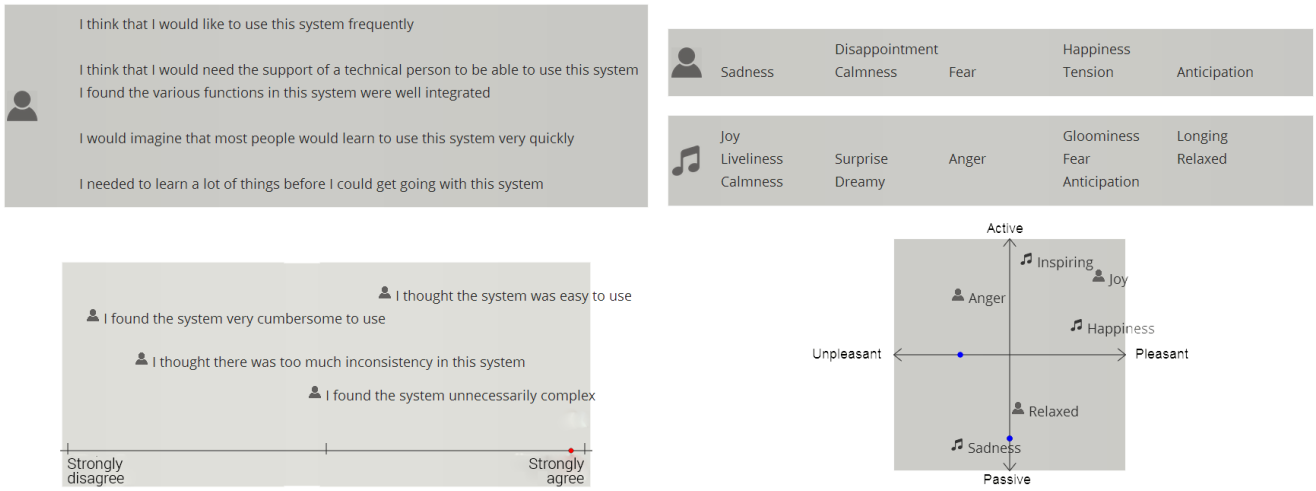


**Figure 1. An example of the MoodStripe (left) and the MoodGraph (right) interfaces. The MoodStripe interface provides a one-dimensional continuous canvas and the item container (holding set of labels). The extended version of the MoodStripe interface, the MoodGraph, has a two-dimensional canvas and a two-category container (here, as used by Pesek et al.).**

based on usability test of the GMail (http://www.gmail.com) product.

The paper is structured as follows: the MoodStripe interface is presented in Section 2, the evaluation is outlined in Section 3 and the results are presented in Section 4. Finally, we conclude the paper in Section 5.

## THE MOODSTRIPE

The novel visual interface, the MoodStripe, was developed with an intention to provide dynamic and intuitive continuous-scale alternative to the standard *n*-degree scale interfaces. The MoodStripe can be easily extended to more complex variations of the basic interface, for example to provide several categories of labels, or additional dimensions, as shown in an example of the MoodGraph interface in Figure 1. Both interfaces are a result of an existing research on music perception and cognition, and were successfully applied to gather user feedback on emotional and visual qualities of music [8].

The MoodStripe interface (Figure 1, left), is implemented as a canvas with one horizontal dimension. The dimension represents the presence of a variable, ranging between two extremes (e.g. negative/positive, absent/significantly expressed, completely disagree/agree). The user is provided with a set of labels, describing different nominal values of the variable. By dragging the labels onto different positions of the canvas, the user marks her perception of each individual label on continuous scale. Positions of placed items can subsequently be quantized to discrete values, if so desired for analysis. The amount of information retrieved by the MoodStripe interface is therefore at least equal to the amount of information gathered by a matrix of radio buttons (for example, a set of 5-degree scales) commonly used to capture similar information.

## EXTENDING THE MOODSTRIPE

The interface can be extended in several ways: by adding categories of variables and/or by adding additional dimensions to the canvas. The MoodGraph interface is an example both such extensions. The MoodGraph interface (Fig. 1, right) integrates two tasks: obtaining results of several categories in one interface, and positioning of selected items onto a two-dimensional space. An alternative to the MoodGraph would be to implement the tasks with a set of two radio groups or check boxes for both categories, and a canvas to position each item onto the two-dimensional space. However, the MoodGraph interface significantly reduces the complexity as it enables us to merge all tasks into one compact form. By dragging the items provided in the category containers (displayed above the canvas) onto the canvas, an icon representing the category appears next to the dragged word (see Figure 1). By using the proposed approach, users can spatially order or reorder the items and visually compare their values (positions) on the same canvas. Consequently, their feedback should be more relevant as the relative comparison of positions helps them decide on individual values. The interface can be further customized by adding rules; for example, that at least one item from each category should be used.

## EVALUATION

The MoodStripe and MoodGraph interfaces were previously used for gathering user feedback on the relations between music, emotions and colors [8]. For that purpose, a preliminary evaluation using a modified version of NASA TLX questionnaire has been conducted [7]. However, in order to fully evaluate the potential of proposed interfaces, also with an intention of possible applications in other domains, we carried out a cross evaluation of the MoodStripe interface using the SUS questionnaire. Additionally, we performed score comparison in order to estimate the differences and possible noise caused by the proposed interface.

The experiment was performed as follows: each participant was asked to evaluate the usability of the GMail product by selecting her answer on the 5-degree scale, as proposed by the

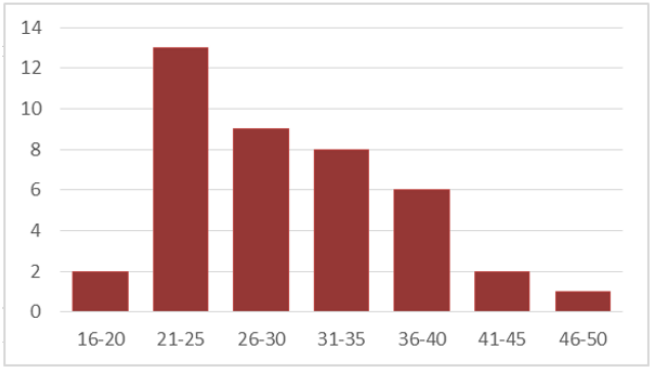


**Figure 3. The distribution of participants according to their age (merged into 5-year groups).**

creators of the original SUS questionnaire (Figure 2, left). Simultaneously, each participant was asked to perform the same task by positioning identical SUS statements in the MoodStripe interface (Figure 2, right). Both tasks were presented in a randomized order to alleviate the question-order bias (balanced study). Finally, each participant was presented with three additional questions, this time providing feedback on the 7-degree scale:

- By comparing both, the MoodStripe and the 5-degree scale interfaces, which of the interfaces was more intuitive and comprehensible? (1 — 5-degree scale, 7 — MoodStripe)
- By comparing both, the MoodStripe and the 5-degree scale interfaces, which of the interfaces takes more time to fill-in? (1 — 5-degree scale, 7 — MoodStripe)
- Do you find that by using the MoodStripe interface, you can express your opinion more easily or more difficultly (due to the visual comparison of your answers to the questions)? (1 — more easily, 7 — more difficultly)

The questionnaire was presented to the participants in Slovenian language, using an evaluated translation of the SUS questionnaire by [3]. Basic demographic data (age and gender) was gathered about each participant and they were also able to provide an optional feedback on the survey (limited to 1000 characters).

## RESULTS

We collected feedback from 41 participants, 12 were male and 29 female. The Figure 3 shows the distribution of the participants according to their age (in 5-year groups) with an average of 29.4 years and standard deviation of 7.1 years.

The scores of the SUS questionnaire are shown in Table 1. The results indicate consistent responses gathered by both interfaces. However, the standard deviation of the responses gathered by the MoodStripe interface is smaller. This is due to the continuous scale of the interface, which allows for a more fluid positioning of the labels, unlike restricted options on traditional n-degree scale. We performed the binarization of responses gathered by the MoodStripe interface into a 5-degree scale and results show comparable scores.

Here, we provide an observation of the responses gahered by the three additional questions described in Section 3. The distributions of responses are shown in Figures 4, 5 and 6 respectively. The results favor the MoodStripe interface over the 5-degree scale: mostly in terms of intuitiveness, less in terms of simplicity of expressing an opinion, and least in terms of time needed for filling out the interface. The latter was somehow expected due to the relatively time-consuming drag-and-drop actions needed for filling-in the responses in the MoodStripe interface, and relatively compact (short) structure of the survey, where the size of the radio button matrix has been kept to a minimum. However, the results confirm previous evaluation of the interface and our overall assumption about the intuitiveness of the proposed MoodStripe interface.

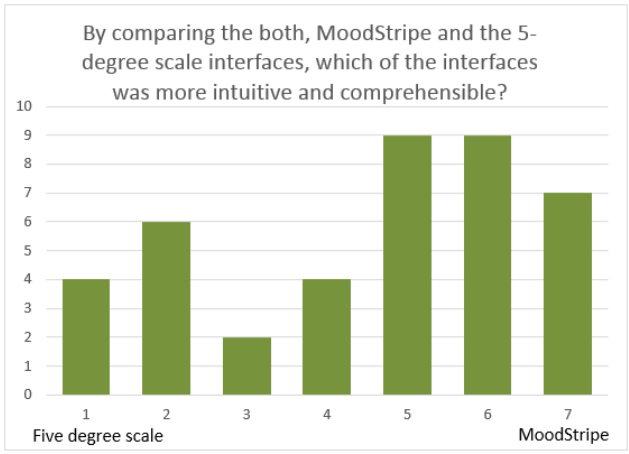


**Figure 4. The responses to the first question. Result show the MoodStripe to be more intuitive than the five degree scale.**

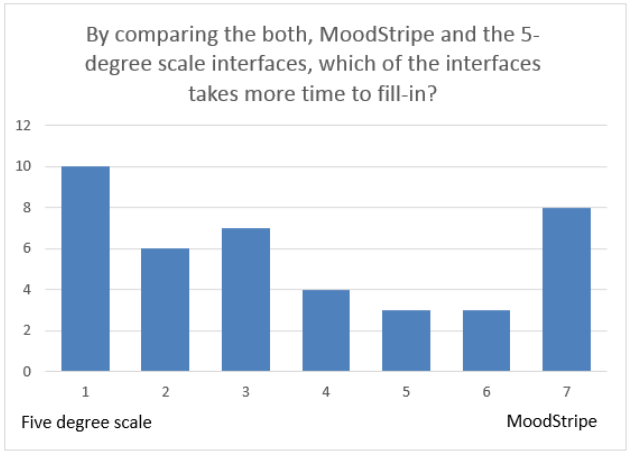


**Figure 5. The responses to the second question. Results lean towards the belief the 5-degree scale takes more time to fill in.**

## CONCLUSION

The paper presented an innovative MoodStripe interface and the results of user experience evaluation. The proposed interface can be applied to a variety of domains and has been previously applied to the domain of music information retrieval. We studied the possibility of applying the visual interface as

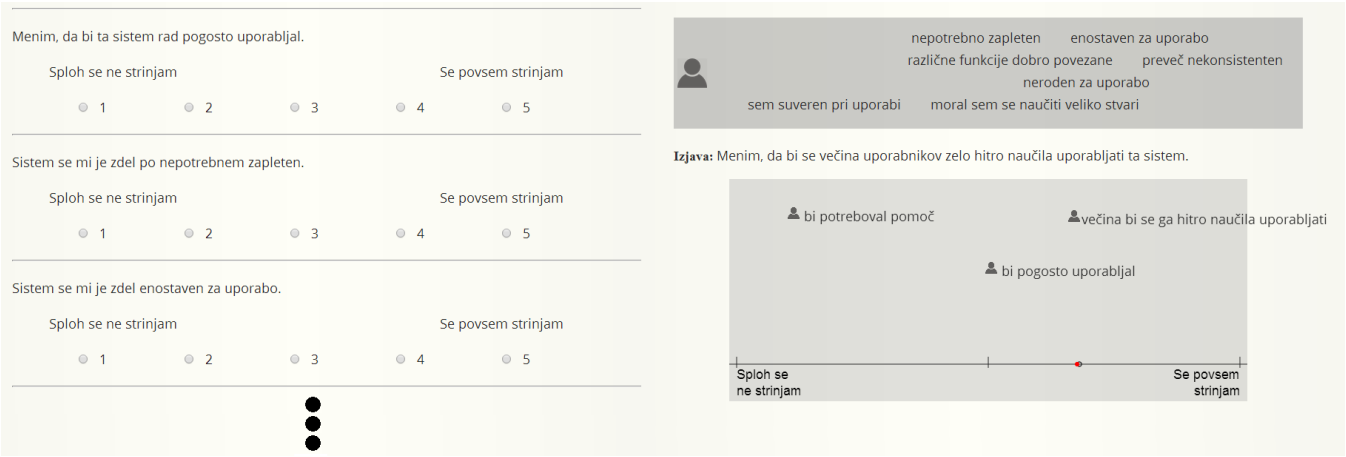


**Figure 2. The evaluation questionnaire. The left side shows a part of the standard SUS questionnaire in the form of traditional matrix of radio buttons. The right side shows the dynamic MoodStripe interface used in the experiment. The statements are shortened into phrases for better overview of the label set, but the full statement for each label is shown on 'mouse-over', below the category container.**

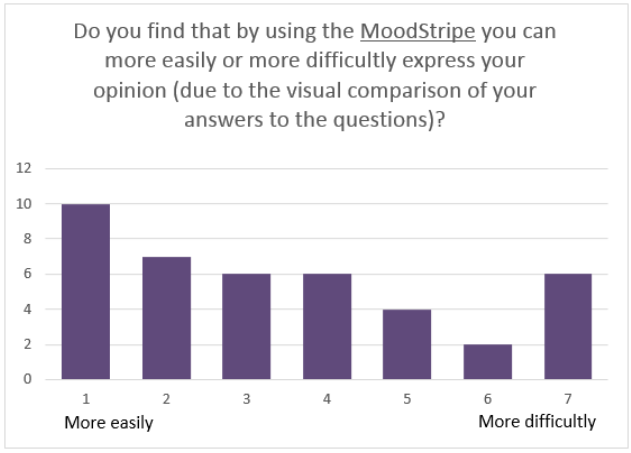


**Figure 6. The responses to the third question. Result show the participants favour the MoodStripe in terms of the simplicity of expressing their opinion.**

a substitute to the traditional *n*-degree scales in the system usability score questionnaire. We compared the obtained SUS scores of both approaches. Additionally, we evaluated the intuitiveness, time consumption and ease of expressing an opinion in both approaches. The results show the MoodStripe interface provides the same quality of the gathered feedbacks, while offering a better user experience. The key advantages of the proposed MoodStripe interface, compared to the standard input types, are the increased intuitiveness and simplicity of expressing opinion by using the interfaces, possibly resulting in lower mental difficulty and frustration of participants.

We have also elaborated on a detailed plan of evaluation and intend to fully develop the interfaces as a strong alternative to standard widgets, such as radio buttons and sliders. Since there is no fixed standard for user information gathering in the field of music information retrieval, as presented by [8], it is our goal to create such standard with the proposed interfaces. We also intend to deliver a tool-kit based on the *JQuery* library in order to further popularise the interfaces.

**Table 1. The average SUS scores and standard deviations. Results indicate consistent responses gathered by both interfaces. Deltas indicate the difference between the questionnaires (marked 1 through 3) for the average and the standard deviation of scores respectively.**

| | avg. SUS score | $\sigma$ of SUS scores |
|---|---|---|
| 1. 5-degree SUS | 79.88 | 18.03 |
| 2. MoodStripe SUS | 79.02 | 16.61 |
| 3. Bin. MoodStripe | 80,55 | 17.27 |
| $\Delta$ 1 vs. 2 | 0.86 | 1.42 |
| $\Delta$ 1 vs. 3 | 1.67 | 0.76 |

### Further plans on evaluation

We plan on developing an extended study for comparison of two products using one interface. For this purpose, we will use the two-category MoodGraph, with each category representing one product. In order to exploit the advantages of two-dimensional representation in the MoodGraph, we plan to simultaneously evaluate several combinations of evaluated software features (e.g. usability, user experience, reliability etc.). The results obtained by such approach will be further compared to a standard test (with the capability of product comparison), such as AttrakDiff. Moreover, we plan on joining two same-type questionnaires (e.g. SUS on two products) into one questionnaire, replacing the questions with MoodGraph where possible.

It is our intention to improve the user experience during the evaluation procedures, and at the same time assess whether the proposed interfaces induce less noise compared to standard interfaces. We plan on evaluating this feature by a large-scale comparison of the user feedback data between the proposed and standard interface-types.

Moreover, the final goal is to provide an interface with minimal noise, induced by the questionnaire interface. Perhaps the greatest challenge is to strictly define the *ground truth* of the responses in this case. This is therefore the reason we initially compared our study to the standard Likert scale for

gathering user feedback. Thus, the challenge of minimal induced noise still remains unsolved.

## REFERENCES


1. Albert, W., and Tullis, T. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Google eBook)*. Newnes, 2013.
2. Bangor, A., Kortum, P., and Miller, J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies 4*, 3 (2009).
3. Blažica, B., and Lewis, J. R. A Slovene Translation of the System Usability Scale: The SUS-SI. *International Journal of Human-Computer Interaction In press* (2014).
4. Hart, S. G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50*, 9 (Oct. 2006), 904–908.
5. Hassenzahl, M., Burmester, M., and Koller, F. AttrakDiff: A questionnaire to measure perceived hedonic and pragmatic quality. *Mensch & Computer* (2003).
6. Lewis, J. R. Critical Review of 'The Usability Metric for User Experience'. *Interacting with computers 25*, 4 (2013), 320–324.
7. Pesek, M., Godec, P., Poredoš, M., Strle, G., Guna, J., Stojmenova, E., Pogačnik, M., and Marolt, M. Capturing the Mood: Evaluation of the MoodStripe and MoodGraph Interfaces. In *Management Information Systems in Multimedia Art, Education, Entertainment, and Culture (MIS-MEDIA), IEEE Internation Conference on Multimedia & Expo (ICME)* (2014), 1–4.
8. Pesek, M., Godec, P., Poredoš, M., Strle, G., Guna, J., Stojmenova, E., Pogačnik, M., and Marolt, M. Gathering a dataset of multi-modal mood-dependent perceptual responses to music. In *Proceedings of the EMPIRE workshop* (2014).